Estimating Mean Cumulative Functions from Truncated Automotive Warranty Data

Jeff Robinson
General Motors R&D Center
Warren, Michigan, USA
jeffrey.a.robinson@gm.com

Stefanka Chukova
Victoria University of Wellington
Wellington, NZ
stefanka@mcs.vuw.ac.nz

Abstract

This article deals with a type of truncation that occurs with typical automotive warranties. Warranty coverage and the resulting claims data are limited by mileage as well as age. Age is known for all sold vehicles all the time, but mileage is only observed for a vehicle with a claim and only at the time of the claim. Here we concentrate on univariate solutions taking either age or mileage as the usage measure. We take a nonparametric approach, so the methods are extensions of the usual calculations for the mean cumulative number of claims or cost of claims and its standard error. We illustrate with real data on four cases based on whether the usage measure is age or miles and whether the results are adjusted for withdrawals from warranty coverage.

1 Introduction

Data from warranty systems are of interest to manufacturers for several reasons. Paid claims are a cost of doing business and a liability incurred by the manufacturer at the time of sale. For these reasons forecasting warranty expenses is of interest. Also, warranty data provide information about the durability of products in the field and, therefore, is of interest to engineers. Furthermore, warranty coverage can be regarded as a product attribute that affects buying decisions. Manufacturers may wish to change warranty coverage to attract more buyers, and they would want to estimate how much the changes would cost. See Robinson and McDonald (1991) for more discussion on these points. Here we emphasize automotive warranties as they are typically done in the U.S.A. The warranty guarantees free repairs subject to both age and mileage limits. The most common limit is now thirty-six months or thirty-six thousand miles, whichever comes first. Age is known all the time for all sold vehicles because sales records are retained. But odometer readings are only collected in the dealership at the time of a claim. So automotive warranties are a case where two usage measures are of interest and information on one of them is incomplete.

The warranty literature is vast and we make no attempt at completeness here. Automobiles are repairable systems, and warranty claims can be thought of as recurrent events associated with the system. We take this approach although for a small enough range of warranty labor operations, the likelihood of duplicate claims on the same vehicle is small enough that methods based on lifetimes (time to first warranty claim) may be appropriate. We also prefer a nonparametric approach because sample sizes are large. We wish to consider two usage measures, one of which is incompletely observed. Finally, we want to deal explicitly with the fact that the events are coming from a warranty plan with specific restrictions, i.e., repair events that may occur outside the limitations of the warranty plan will not be included in the database. Even considering these restrictions there is some relevant past literature. The model and estimation procedure we discuss later is an extension of the "robust estimator" presented by Hu and Lawless (1996b). Their approach extends the typical Nelson-Aalen estimator (Nelson 2003) to account for, in our case, the probability of a vehicle not being eligible to generate a warranty claim because it has exceeded its warranty coverage limits. Lawless, Hu and Cao (1995) also explicitly deal with our censoring problem and specify a simple linear automotive mileage accumulation model, which we use here to estimate the current mileage of a vehicle for which we have a previous odometer reading. They also present a family of semi-parametric models to relate event times to mileage accumulation. Lawless (1998) presents a survey and some extensions including dealing with the bias caused by the reporting delay of warranty claims. This issue was also previously discussed in Lawless and Nadeau (1995) and in Kalbfleisch, Lawless and Robinson (1991). The bias issue is of interest here because using the warranty data itself to estimate mileage accumulation rates can create a form of bias as well. A finite sample population correction to the variance estimate is discussed in Robinson (1995). In addition to their other work mentioned, Hu and Lawless (1996a) also developed a general framework for handling the type censoring considered here making use of supplementary information. Relevant methods applicable to warranty are also found in Blischke and Murthy (1996).

2 The Model and Estimation Procedure

We utilize notation from Hu and Lawless (1996b). Let $n_i(t)$ be the number of claims (or cost associated with those claims) at time t for vehicle i. It will be convenient and not restrictive to think of time as discrete, i.e., $t = 1, 2, ...$ Let $N_i(t)$ be the accumulated number of claims (or cost) up through and including time t. "Time" here will be either age of the vehicle or mileage, not calendar time. Suppose M such vehicles have been under observation, i.e., their records are part of the warranty database. Let τ_i be the "time" that unit i has been under observation. The exact definition of the τ's will depend on whether time is age or mileage. We will certainly be interested in the population cumulative mean function $\Lambda(t) = EN_i(t)$ for various t. Sometimes it will be convenient to think of M as a sample from a larger population of vehicles, some of which have not yet been sold. Even after they have all been sold, we want to estimate $\Lambda(t)$ for values of t greater than we have observed for some of the vehicles. We will also want to put a standard error on the estimate. For the discrete time case, the incremental rate function is $\lambda(t) = \Lambda(t) - \Lambda(t-1)$, where $\Lambda(0) = 0$. Let $\delta_i(t) = I(\tau_i \geq t)$ be the indicator of whether car i is under observation at time t. Then $n_.(t) = \sum_{i=1}^{M} \delta_i(t) n_i(t)$ is the total number of claims (or cost) observed at time t for all M vehicles. If the observation times are independent of the event process, the rate function is estimated by

$$\hat{\lambda}(t) = \frac{n_.(t)}{MP(t)},$$

where $P(t)$ is the probability that a vehicle of age t is eligible to generate a claim. This is the "robust estimator" stated in Hu and Lawless (1996b). Their $P(t)$ is a right-hand tail probability of an age distribution and it is assumed known. These assumptions are relaxed here. The associated cumulative mean function estimator is $\hat{\Lambda}(t) = \sum_{s=1}^{t} \hat{\lambda}(s)$ for $t = 1, 2, \ldots, \max_{1 \leq i \leq M}(\tau_i)$. Hu and Lawless show asymptotic normality under mild conditions with a standard error for $\hat{\Lambda}(t)$ given by the square root of

$$\hat{V}ar[\hat{\Lambda}(t)] = \frac{1}{M^2} \sum_{i=1}^{M} \left(\sum_{s=1}^{t} \left[\frac{\delta_i(s) n_i(s)}{P(s)} - \hat{\lambda}(s) \right] \right)^2 .$$

Of course in practice the probability function $P(t)$ is not known. We will estimate it from the same warranty data that we use to estimate the rate function, and by doing so ignore two facts. First, the uncertainty associated with the estimate of $P(t)$ is ignored. Second, our estimate of $P(t)$, which in our case will be based on estimated mileage accumulations, may be biased because of the way warranty data accrues in practice. For example if warranty claims occur primarily due to mileage accumulation, then early in the model year high-mileage vehicles may be over represented in the set of observed claims. We will attempt to deal with these issues as best we can by utilizing the size and continuing nature of the warranty database.

3 Example

First, we consider "time" to be the age of the vehicle. If we ignore mileage entirely, then τ_i is simply the current age of the vehicle in days. The probability $P(t)$ is estimated by the proportion of vehicles with ages greater than t, i.e., $\hat{P}(t) = \frac{\#(\tau_i \geq t)}{M}$, and the estimator above is the "standard" Nelson-Aalen estimator.

This estimate correctly characterizes how warranty expense accrue with respect to age ignoring miles. It may be useful for prediction particularly if mileage accumulation rates remain stable. But it does not reflect

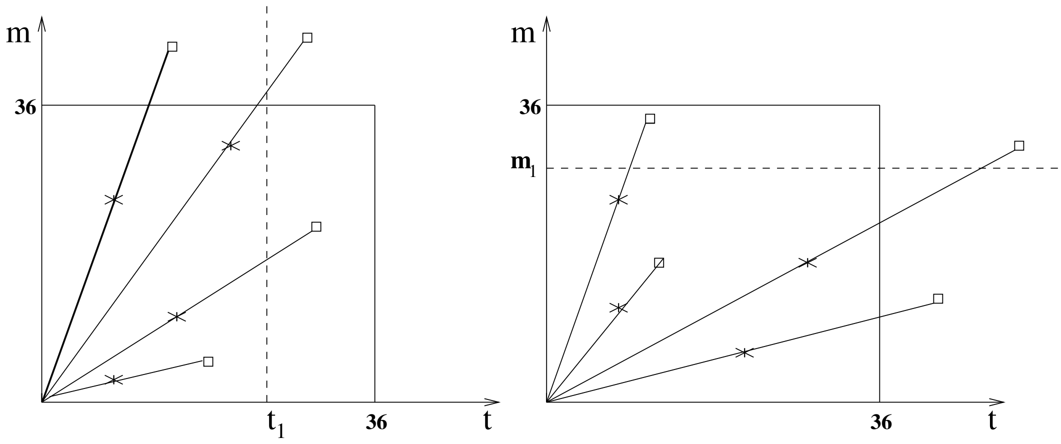

Figure 1: Censoring situations for “time”= age (left) and “time” = miles

the ”true” rate per vehicle because it does not reflect the elimination of vehicles eligible to generate claims due to the mileage limit. To capture this rate we define $\tau_i = \min(a_i, y_i)$, where y_i is the age in days at which the vehicle exceeds the mileage limit of 36,000 miles and a_i is its current age. Since odometers are not monitored continuously, y_i is not known. For a vehicle that has had at least one claim, we use the estimate $\hat{y}_i = \frac{36{,}000 c_i}{m_i}$, where c_i and m_i are the age and mileage of the latest claim. Let m be the number of vehicles with at least one claim. The proportion of such vehicles that are estimated to be within warranty coverage at age t is given by

$$\hat{P}(t) = \frac{1}{m} \# \left(\frac{m_i}{c_i} \leq \frac{36,000}{t} \right).$$

We take this as the estimate for all vehicles including those that have had no claims.

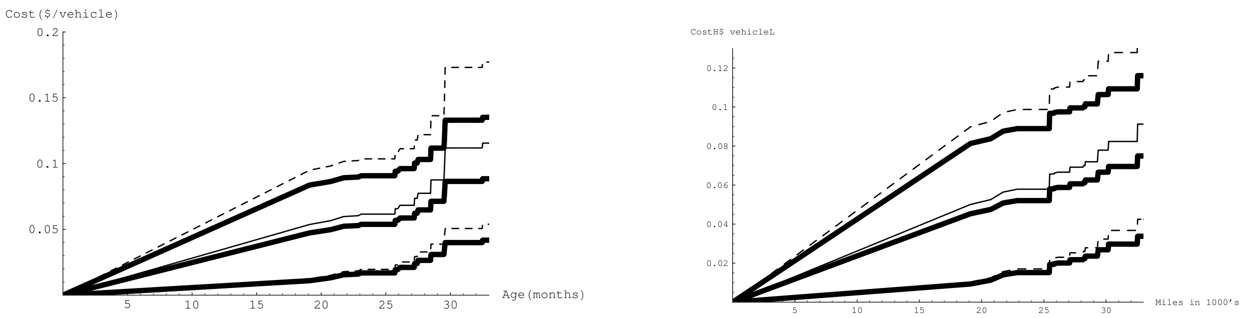

Figure 2: Adjusted and unadjusted (in bold) estimates for $\Lambda(t)$ with “time”= age (left) and “time” = miles

The various censoring situations are displayed in Figure 1, which shows four typical cars. The values t_1 and m_1 represent a target age and mileage respectively. The “ $*$ “ represents the observed claims, and “$\square$“ represents the estimated current age and mileage for each vehicle. On the left-hand plot, even if mileage is ignored, only the middle two lines are eligible to generate a claim at the target age. But restricting for mileage the top one of these becomes ineligible because it is estimated to have left warranty coverage prior to the target age. Similarly on the right-hand plot, only the first and third vehicles (proceeding clockwise) would be eligible at the target mileage. The lower one of these would be eliminated by restricting for age. The results of applying this methodology are shown in Figure 2 for the cost of claims associated with a subsystem of the vehicle. For convenience the age axis is displayed in months (30 day increments). There were 752 vehicles in the dataset ranging in age mostly from 24 to 36 months. Plots are shown for the mean function and plus/minus two standard errors. As expected the mean function for the adjusted cases are shifted upward to reflect vehicles exiting warranty coverage.

4 Discussion

We have adopted a very pragmatic approach here and ignored a few important considerations such as the bias introduced by estimating mileage accumulation rates from the warranty data itself. In the full paper we explore more fully the impact of taking mileage information from warranty records from previous model years. We are particularly interested in controlling for bias in the mean function estimate because we have found that such biases can lead to misinterpretation in practice and reluctance to use the methods. We also consider using the "restricted" estimates discussed earlier to estimate the effects of making small changes to warranty coverage.

References

Blischke, W.R. and Murthy, D.N.P. (1996). *Product Warranty Handbook*. Marcel Dekker.

Hu, X. and Lawless, J. (1996a). Estimation from truncated lifetime data with supplementary information on covariates and censoring times. *Biometrika 83*, 747–761.

Hu, X. and Lawless, J. (1996b). Estimation of rate and mean function from truncated recurrent event data. *Journal of American Statistical Association 91*, 300–310.

Kalbfleisch, J., Lawless, J., and Robinson, J. (1991). Method for analysis and prediction of warranty claims. *Tecnometrics 33*, 273–285.

Lawless, J. (1998). Statistical analysis of product warranty data. *International Statistical Review 66*, 41–60.

Lawless, J., Hu, J., and Cao, J. (1995). Method for estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis 1*, 227–240.

Lawless, J. and Nadeau, J. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics 37*, 158–168.

Nelson, W. (2003). *Recurrent Events Data Analysis for Products Repairs, Disease Recurrences, and Other Applications*. SIAM.

Robinson, J. (1995). *Recent Advances in Life-Testing and Reliability*, Chapter Standard Errors for the Mean Number of Repairs on Systems from a Finite Populations, pp. 195–217. CRC Press.

Robinson, J. and McDonald, G. (1991). *Data Quality Control: Theory and Pragmatics*, Chapter Issues Related to Field Reliability and Warranty Data. Marcel Dekker.